Challenges

Sentiment Analysis is an umbrella term for large number of emotions and affect related discoveries or tasks, each of which has their own distinct challenges. The sub-sections provide an overview of some of them.

1. Sentiment at different text granularities

These days Sentiment can be determined from various mediums: from sentiment associations of words and phrases; to sentiment of sentences, SMS messages, chat messages, and tweets; to sentiment in product reviews, blog posts, and whole documents. Modals and negation impact affect of the sentence without them having strong sentiment associations. For example*, The war has created millions of refugees.*

Here, the sentence is neutral for a negative situation from the message producer’s perspective.

It is possible that a particular text shows a particular feeling without use of any overt and explicit markers. For example, *Another Monday, another week, working my tail off.* This clearly conveys sense of frustration without use of any highly specific words for showing frustration.

Many times sentiment in sentence is judged on the basis of component words in it but the valence of a sentence is not simply the sum of the polarities of its constituent words.

With increasing number of social media websites, there is also an increasing trend of using non-standard texts.

For example, consider

*misspellings - parlament*

*creatively spelled words - luv hashtagged words - #luvumom*

*abbreviations - tbh, af, f9, lmao*

1. Detecting sentiment of writer, reader and other entities

On just having a look at a sentence, it may seem unambiguous but looked closely, it is easy to identify if the sentiment is of the writer, reader of any other entity. Many times, it is usually unclear if the sentiment in the text is same as the sentiment of the speaker. It may refer to the emotions without explicitly or implicitly referring to the speaker’s view.

For example, consider

*My friend always keeps on complaining how hectic her schedule is.*

The sentiment in the text is definitely negative (frustration and complaint), but it is not of the speaker itself. It is of someone the speaker knows.

Another example, consider

*Bob: The pop star suffered a fatal overdose of heroine.*

This text describes a negative event (death of a person), but it cannot be concluded that the speaker, Bob, is personally saddened by this negative event. It is possible that Bob is a news reporter and merely describing the event.

1. Outlook Detection

By this term, it means that it should be identifiable from text whether the speaker is in favor, against or neutral towards the subject of the text.

For example, consider

*Congratulations! You broke my $1000 glass!*

The speaker is not actually happy and neither is the speaker congratulating; the tone of the text is sarcastic.

Hence as seen, it is difficult to interpret creative uses of language like sarcasm, irony, humour, metaphor, idiomatic and figurative language.

New research in fields like social media texts, and approaches that combine traditional sentiment analysis with relation extraction can significantly impact in improving the current form of automatic outlook detection.

1. Detecting affect and emotions

Most machine learning algorithms for affect analysis require large amounts of training data because there are numerous affect categories like happiness, sadness, excitement, disgust, frustration etc. Moreover, these affect categories can be overlapping.

For example, consider

*Wow! I am so happy for you.*

This text can be categorized in happiness as well as excitement category.

Moreover, to get a good F-score and higher accuracy, it is important to have huge amounts of training data.

Many a times, the text can have different affect depending upon its sense and context.

For example, consider,

*Mary hugged her daughter before going to work.*

This text is quite emotional whereas,

*The pipeline hugged the state border.* This text is rather unemotional.

But both of the texts have the presence of the same word, that is, ‘hugged’. Therefore, this is for sure that the affect and emotion of the text cannot be decided alone on the polarities of words occurring in it.

1. Named entity recognition and anaphora resolution

*Liar, Liar is a good watch.*

In this text, the speaker is not actually talking about he being a liar or someone else being a liar and neither is he saying that being a liar is a good thing; he is simply talking about the movie ‘Liar, Liar’.

The classifier has to be trained to recognize named entities like movies, places, idioms etc.

It should also be able to resolve what a pronoun, or a noun phrase refers to.

*We watched the movie and went to dinner; it was awful.*

In this text, the classifier should be able to clearly distinguish if it refers to the movie or the dinner.

Researchers are trying to address these challenges and come up with solutions to build a better classifier in each iteration. We will try to analyze these challenges in later section.

Critical Analysis

Challenges